

This application is submitted in the name of inventors Edward Archibald and  
Spencer Brewer.

SPECIFICATION

METHOD AND APPARATUS FOR  
A MIXED-MEDIA MESSAGING SYSTEM

BACKGROUND OF THE INVENTION

Field of the Invention

10        The present invention relates to data communications. In particular, the present  
invention relates to creating a mixed media message.

The Prior Art

Background

The widespread acceptance and use of the Internet has generated much excitement, particularly among those who see the Internet as an opportunity to develop new avenues for communication. Many different types of communications are available over the Internet today, including email, IP telephony, teleconferencing and the like.

One application of the Internet that has received attention is adding multimedia capabilities to traditional email services. For example, such a system may provide a downloadable application which allows the user to record a voice message and send it as an email attachment. The email recipient then receives an email with an MP3-encoded audio file attached to it which can then be played with a standard media player. Some systems allow users to utilize a telephone and add a voice message that will be delivered along with the greeting. Still other systems allow the user to include an image, to record audio and to mix an existing audio file with the recorded audio.

While these systems perform their intended functions, they suffer from certain disadvantages. For example, in systems of the prior art, the mix is 'flat', that is, the

user's recorded message and the audio file are mixed at 'full volume' for their entire length. Therefore, the resulting mixed audio file will not have a "professional" or polished sound and may result in the user's message being obscured by the background track.

5 Another disadvantage of prior art systems is that the audio file being used for the background mix will be included in its entirety. Hence, if the background file is 3 minutes long and the voice file is 10 seconds, the entire mix will be three minutes. Again, users will not perceive such a mix as professional and may be additionally frustrated by the time and bandwidth necessary to download unnecessary audio.

10 Hence, there exists a need to provide a mixed-message system which provides a professional-sounding product wherein the volume levels and length of the message are automatically mixed for the user by the system.

BRIEF DESCRIPTION OF THE INVENTION

The present invention relates to data communications. In particular, the present invention relates to creating a mixed media message.

A method for creating a mixed media message on a client node coupled to a host node over a network such as the Internet is disclosed. The method comprises choosing a soundscape; recording a message; and mixing the soundscape and the message in a predetermined manner. A host node is disclosed which is configured to provide a client node with the means for performing the method is disclosed. A client node is disclosed which is configured to receive means for performing the method is disclosed.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

Figure 1 is a diagram of a mixed-messaging system according to the present invention.

Figure 2 is a flowchart of a method of creating a mixed message according to the present invention.

Figure 3 is a diagram of a soundscape according to the present invention.

Figure 4 is a screenshot of the elements of a user interface (UI) according to the present invention.

Figure 5 is a screenshot of the UI in the soundscape selection phase according to the present invention.

Figure 6 is a screenshot of the UI in the message recording phase according to the present invention.

Figure 7 is a screenshot of the UI in the mix preview phase according to the present invention.

Figure 8 is a flowchart of mixing according to the present invention.

Figure 9 is a diagram of a mixing template according to the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Persons of ordinary skill in the art will realize that the following description of the present invention is illustrative only and not in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons having the benefit of this disclosure.

It is contemplated that the present invention may be embodied in various computer and machine readable data structures. Furthermore, it is contemplated that data structures embodying the present invention will be transmitted across computer and machine-readable media, and through communications systems by use of standard protocols such as those used to enable the Internet and other computer networking standards.

The invention further relates to machine-readable media on which are stored embodiments of the present invention. It is contemplated that any media suitable for storing instructions related to the present invention is within the scope of the present

invention. By way of example, such media may take the form of magnetic, optical, or semiconductor media.

The present invention may be described through the use of flowcharts. Often, a single instance of an embodiment of the present invention will be shown. As is

5 appreciated by those of ordinary skill in the art, however, the protocols, processes, and procedures described herein may be repeated continuously or as often as necessary to satisfy the needs described herein. Accordingly, the representation of the present invention through the use of flowcharts should not be used to limit the scope of the present invention.

10 The present invention may also be described through the use of web pages in which embodiments of the present invention may be viewed and manipulated. It is contemplated that such web pages may be programmed with web page creation programs using languages standard in the art such as HTML or XML. It is also contemplated that the web pages described herein may be viewed and manipulated with web browsers

15 running on operating systems standard in the art, such as the Microsoft Windows® and

Macintosh® versions of Internet Explorer® and Netscape®. Furthermore, it is contemplated that the functions performed by the various webpages described herein may be implemented through the use of standard programming languages such a Java® and similar languages.

5           The present invention will first be described through a diagram which illustrates the structure of the present invention, and then through figures which illustrate the operation of the present invention.

FIG. 1 is a diagram of a mixed-message system 100 according to the present invention. System 100 includes a host system 102 configured to host the present invention. It is contemplated that host 102 may comprise a server configured to run an operating system standard in the art such as Microsoft Windows®, Microsoft Windows® NT, or UNIX®. It is contemplated that host 102 may comprise a single server, or a plurality of servers operatively coupled over a network standard in the art.

Host 102 further includes an application server 104 configured to operate  
15   according to the present invention in a manner described in more detail below. Host 102



further includes a database 105 standard in the art for storing programs and media utilized in the present invention.

Host 102 further includes a web server 106 operatively configured to host a website. Web server 106 may comprise hardware and software standard in the art, and preferably is configured to interpret a language useful in Internet applications, such as JAVA®.

To couple the host 102 to the outside world, typically a gateway 108 standard in the art is provided and operatively coupled between web server 106 and backbone network 110. Backbone network 110 may be any packet-based network standard in the art, such as IP, Frame Relay, or ATM.

To provide additional communications to legacy POTS phone, host 102 may include a Computer-Telephony Integration Service (CTI) 108 configured to provide a telephony user interface (TUI) to users.

The system 100 of FIG. 1 also includes an example of a broadband residential client system 118. To reach individual residences, typically local ISP providers provide a

cable or DSL head end 120 standard in the art. An individual wishing to utilize the ISP's service may then employ a cable modem or DSL modem 122 coupled to the user's home LAN 124. The user's home LAN may be provided by a home PC 126 configured to run software standard in the art such Microsoft Windows®. The user may then operatively  
5 couple an IP telephone 128 to the LAN 124 or directly to computer 126. The user may also operatively couple a microphone 129 directly to computer 126.

FIG. 2 is a flowchart of the creation of a MixedMessage according to the present invention. FIG. 2 provides an excellent overview of the process according to the present invention. In an exemplary non-limiting embodiment of the present invention, the process of FIG. 2 runs on a Windows® platform system including a microphone and speakers. Through the process as shown and described, the present invention creates an audio clip referred to herein as a MixedMessage.

Briefly, the process of FIG. 2 begins in act 200 where a user will select a soundscape. The user then records a message in act 202. Finally, the user mixes the

soundscape and their recorded message together into a MixedMessage and previews the MixedMessage in act 204.

Referring now to FIG. 3, a diagram of a soundscape according to the present invention is shown. The soundscape of FIG. 3 is represented as digital audio data composed of audio content occurring during time periods T1, T2, and T3. In an exemplary non-limiting embodiment of the present invention, time period T1 comprises a front punctuating sound (FPS), time period T2 comprises background audio, and time period T3 comprises a back punctuating sound (BPS). In an exemplary non-limiting embodiment of the present invention a soundscape which is comprised of three segments may be composed of three distinct digital audio files. Further, segments that are used to compose a given soundscape may be employed interchangeably in the composition of other soundscapes. Though the present example shows a three-section soundscape, it is contemplated that soundscapes may comprise a wide variety of content and themes.

By way of example, a soundscape having an ocean theme may comprise a FPS consisting of the surf crashing with the sound of seagulls calling in the distance; the

background may consist of a continuation of the sound of the surf together with a romantic melody being played on an acoustic guitar; and the BPS consisting of the highlighted cry of a lone seagull.

An example of how a user may utilize the present invention over the Internet will

5 now be shown and described.

### The User Interface

Figure 4 is a screenshot of a user interface (UI) 400 according to the present invention. UI 400 may also include a plurality of components having corresponding functions. In an exemplary non-limiting embodiment of the present invention, these components are used throughout the operation of the invention and may have different functions depending on which 'phase' of the user is in.

The invention UI 400 includes a soundscape panel 401. Soundscape panel 401 is enabled to allow a user to select a soundscape. It is contemplated that soundscape panel 401 will conform to file selection standards according to the client node's operating

system. By way of example, soundscape phase 401 is shown operating on a Windows®-compatible personal computer.

UI 400 further includes a phase indicator panel 402. In an exemplary non-limiting embodiment of the present invention, phase indicator panel 402 indicates the user's progress in achieving the steps of the present invention as shown and described in FIG. 2.

UI 400 may also be enabled with media control buttons 403 which control the operation of playback or recording of the present invention depending on the application phase. UI 400 may also have navigation buttons 404 standard in the art which allow the user of the application to move, at appropriate times, between the phases of the application. UI 400 may also include a context sensitive help/status panel 405 standard in the art which allows the user to receive help on the operation of the application and on the current operational status of the application. UI 400 may also include a sound recording/playback progress panel 406 that indicates the current progress of playback or recording as a 'percentage complete' indicator.

UI 400 may also include an image display panel 407 that displays an image corresponding to the selected soundscape.

Choosing a soundscape

The first step of the present invention is to choose a soundscape.

5 FIG. 5 is a screenshot of the invention UI in the soundscape selection phase 500 according to the present invention. It is contemplated that soundscape selection phase 500 will be displayed to a user on a client node which is operatively coupled to a host node which is configured according to the present invention. In an exemplary non-limiting embodiment of the present invention, the invention UI soundscape selection phase 500 and the rest of the phases to follow are generated on a client node using hardware and software standard in the art, such as JAVA and a JAVA®-enabled web browser running on a personal computer.

The soundscape selection phase 500 includes a soundscape panel 501, a phase indicator panel 504, media control buttons 505, navigation buttons 508, context sensitive

help/status panel 507, and a sound recording/playback progress panel 503, which function in a manner substantially similar to that of FIG. 4.

The soundscape selection phase 500 may also include an image display panel 509 that displays an image corresponding to the selected soundscape.

5 As can be seen by inspection of FIG. 5, soundscapes according to the present invention are organized into a plurality of Genre folders within the soundscape panel 501. In an exemplary non-limiting embodiment of the present invention, double-clicking on any of the soundscape Genre folders in the soundscape panel 501 or clicking on the '+' opens the folder to display a further set of folders at an 'Edition' level.

10 In an exemplary non-limiting embodiment of the present invention, when a user has opened an 'Edition' level folder, the user will be presented with one or more available soundscapes as indicated by a speaker icon or other suitable indicator. As can be seen by inspection of FIG. 5, the "Monet's Water Garden" soundscape 502 has been selected. In an exemplary non-limiting embodiment of the present invention, when a  
15 user has selected a given soundscape a corresponding image may displayed in the image

display panel 509. In an exemplary non-limiting embodiment of the present invention, the image may be downloaded from the host web server.

Thus, in the example shown in FIG. 5, the user has chosen the "Monet's Water Garden" soundscape 502 from the "Soothing" Edition of the "Music" Genre.

5 It is important to note that the various soundscapes may be sorted within soundscape selection panel 501 by emotional characteristics or other methods. Additionally, soundscapes may be organized by pictures or other indicators such as icons. Additionally, soundscapes and other sources of sounds may be presented by completely different means independent of such a hierarchical representation.

10 In an exemplary non-limiting embodiment of the present invention, the soundscapes are stored on a media database on the host server and when the user selects a particular soundscape, that soundscape is presented to the client by 'streaming' highly compressed digital audio data to the client to minimize any delay. In an exemplary non-limiting embodiment of the present invention, the soundscape components (i.e. FPS, BG  
15 and BPS) are streamed in MP3 format and are converted, on the client, into raw PCM



data. In an exemplary non-limiting embodiment of the present invention, the soundscape components will be stored in the RAM of the client node and ultimately played for the user.

A user may select a particular soundscape by single-clicking on it and then can  
5 control the playback of the soundscape by using the media control buttons 506. At any  
given time, the soundscape that is highlighted in the soundscape panel becomes the  
soundscape that will be used in other phases of the MixedMessage creation and use.

As can be seen by inspection of FIG. 5, the button labeled 'soundscape' of phase  
indicator panel 504 is highlighted to indicate that the user is in the soundscape selection  
10 phase.

FIG. 5 as shown and described provides an example of means for selecting a  
soundscape.

#### Recording a message

The next step is for the user to record a message.

FIG. 6 is a screenshot of the UI in the recording phase 600 according to the present invention. It is contemplated that the recording phase 600 will be displayed to a user on a client node which is operatively coupled to a host node which is configured according to the present invention.

5 By way of example, recording phase 600 is shown operating on a Windows®-compatible personal computer running a JAVA®-enabled web browser.

The recording phase 600 includes a soundscape panel 607, a phase indicator panel 602, media control buttons 604, navigation buttons 606, context sensitive help/status panel 605, and a sound recording/playback progress panel 601, which all function in a manner substantially similar to that of FIG. 4.

The recording phase 600 may also include an image display panel 608 that displays an image corresponding to the selected soundscape.

To start recording, the user may click the 'record' button 603 in the media control buttons 604. In an exemplary non-limiting embodiment of the present invention,

15 recording commences immediately and continues until the user presses the 'stop' button

in media control button section 604. The user can then audition the recorded message utilizing the play, pause and stop buttons in media control button section 604.

In an exemplary non-limiting embodiment of the present invention, the user's voice is recorded using standard hardware and software on the user's computer. In a presently preferred embodiment, the present invention performs necessary media control functions by interfacing with the user's PC using a protocol such as Direct-X®. In an exemplary non-limiting embodiment of the present invention, the voice data is stored in RAM memory in PCM format.. For permanent storage, the voice data may be stored as a .wav file on the client node.

In an exemplary non-limiting embodiment of the present invention, the user may indicate that they are satisfied with their message and are ready to move on to the next step by pressing the control button 606 or the appropriate application phase indicator button 602.

As can be seen by inspection of FIG. 6, the button labeled 'message' of phase indicator panel 602 is highlighted to indicate that the user is in the record phase.

Additionally, the bar in sound recording/playback progress panel 601 has lengthened to indicate the user's further progress through the present invention.

The recording phase 600 as shown and described provides an example of means for recording a message.

### Mixing

The next step is to mix the chosen soundscape with the recorded message.

FIG. 7 is a screenshot of the invention UI in the mix review phase 700 according to the present invention. It is contemplated that mix review phase 700 will be displayed to a user on a client node which is operatively coupled to a host node which is configured according to the present invention.

By way of example, mix review phase 700 is shown operating on a Windows®-compatible personal computer running a JAVA®-enabled web browser.

The mix review phase 700 includes a soundscape panel 701, a phase indicator panel 706, media control buttons 703, navigation buttons 705, context sensitive

help/status panel 704, and a sound recording/playback progress panel 702, which function in a manner substantially similar to that of FIG. 4.

The mix review phase 700 may also include an image display panel 707 that displays an image corresponding to the selected soundscape.

5 In an exemplary non-limiting embodiment of the present invention, the user initiates the mixing process by clicking on the 'play' button in media control buttons 703. The present invention then mixes the recorded message and the chosen soundscape in a predetermined manner. In an exemplary non-limiting embodiment of the present invention, the soundscape is streamed from the server, decoded, and stored in the client's 10 RAM; the recorded message is then read from the client node's RAM; and all of the aforementioned components are then mixed in a predetermined manner and immediately played for the end user as audio. The progress of the mixing process may be displayed to the user through progress indicator 702.

In preferred embodiments, the actual mix takes place in real time and the end user 15 hears the result immediately. By performing the mixing process on the client node, the

present invention allows the mixing process to occur in real time. This immediacy of the end-user feedback is a significant improvement over systems of the prior art and provides users with increased convenience. For example, in systems utilizing the present invention, users may then choose different soundscapes with their recorded message, and

5 hear the preview immediately.

In an exemplary non-limiting embodiment of the present invention, the recorded voice message is processed using audio processing tools standard in the art prior to the mixing process.

As can be seen by inspection of FIG. 7, the button labeled 'mix it!' of phase indicator panel 706 is highlighted to indicate that the user is in the mix and review phase.

The mixing phase 700 as shown and described provides an example of means for mixing and reviewing a message.

FIG. 8 is a flowchart of a method for mixing a soundscape and a recorded message according to the present invention. The process of FIG. 8 may be performed by media

15 tools standard in the art running on the client node. In an exemplary non-limiting

embodiment of the present invention, the length of the recorded message may also be determined for use with other aspects of the present invention.

The process of FIG. 8 begins with the present invention examining the recorded message and removing any silence from the beginning and end of the recorded message in act 800. In an exemplary non-limiting embodiment of the present invention, act 800 is performed using audio gating tools standard in the art.

The process of FIG. 8 continues with the recorded message being normalized in act 802. As is appreciated by those of ordinary skill in the art, the process of normalization involves the use of compression and limiting tools which adjust the peak and average levels of an audio signal in a predetermined manner, thus giving an audio signal a "studio" quality. In an exemplary non-limiting embodiment of the present invention, act 802 may be performed using compression/limiter tools standard in the art.

Tim: We intend to do more processing than is indicated here.

The process of FIG. 8 concludes with act 804, where the present invention interleaves the gated and normalized recorded message with the chosen soundscape in a predetermined manner.

It is contemplated that other acts may be performed during the mixing process in addition to those listed in FIG. 8. For example, additional process steps such as trimming or harmonizing may be performed on the recorded message; the length of the voice track may be calculated; the mix contours for all tracks may be calculated, such as the starting offset, volume changes, ramp up and ramp down time, and ending offset as will be further described below. Additionally, the present invention may initiate streaming of the user's recorded message back to the host server, or output the result of the mix to a digital audio file on the client node

FIG. 9 is a diagram of an interleaving process according to the present invention.

FIG. 9 demonstrates how the various audio tracks which comprise a MixedMessage are assembled into a final product. FIG. 9 shows each individual track with its corresponding



level contour all displayed along a common time axis, with particular time intervals of interest numerically indicated along a top time axis.

FIG. 9 shows one aspect of the present invention where a front punctuating sound, a background, a recorded message, and a back punctuating sound are interleaved together to form a MixedMessage. As can be seen by inspection of FIG. 9, each individual track has a corresponding contour, where each contour has a corresponding level and time offset. The overall effect of the tracks and their corresponding contours may be said to form a MixedMessage template.

It is contemplated that a MixedMessage template may comprise any number of individual tracks. It is further contemplated that each individual track may consist of any multimedia information suitable for display or presentation to a user. Though the present example consists of audio information be interleaved, it is contemplated that each audio track may consist of a sub-mix of audio information mixed in a previous mixing process. It is further contemplated that other media, such a video information, may be included in the process of the present invention.

Referring back to FIG. 9, and more specifically to track 1, the front punctuating sound (FPS) is first mixed in. In an exemplary non-limiting embodiment of the present invention, the FPS is mixed in by bringing the FPS into the mix by raising its level at a predetermined rise time during time interval 1. The level of the FPS is then held for a predetermined length of time during time interval 2, then the FPS is removed from the mix by dropping its level at a predetermined fall time during time interval 3. In a presently preferred embodiment, time interval 1 lasts approximately 0.1 seconds, time interval 2 lasts approximately 1 to 3 seconds, and time interval 3 lasts approximately 0.1 seconds.

Referring now to track 2, the background is mixed in a manner similar to the FPS during time intervals 1-4. However, during time interval 5, the background level is lowered to a predetermined level at time interval 5. In a presently preferred embodiment, the background is lowered to a non-zero level until time interval 8, referred to as the bed volume. In an exemplary non-limiting embodiment of the present invention, the bed volume is approximately 10 - 18 dB below the level of the recorded message. Referring now to track 3, the recorded message is mixed in during time interval 6 by raising the

level of the message at a predetermined rise time. The message is then played for its predetermined length during time interval 7. The message is then removed from the mix by lowering its level at a predetermined fall time during time interval 8.

After the message has concluded in time interval 8, the background level is then raised at a predetermined rise time during time interval 9, and then may be played for a predetermined amount of time during time interval 10.

Referring now to track 4, the back punctuating sound (BPS) may be brought into the mix by raising its level at a predetermined rise time during time interval 11. The BPS level may then be maintained for a predetermined amount of time during time interval 12. Finally, to conclude the MixedMessage, both the BPS and the background may be mixed down by lowering their levels at a predetermined fall time during time interval 13.

As can be seen by inspection of FIG. 9, the template according to the present invention may be characterized as mixing predetermined information around a message provided by a user.

After the processes described above are complete, additional acts may be performed. For example, when the user is satisfied with the result, an upload phase may

This XML document is then sent to the MixedMessages Application Server and represents instructions by which the host node may reproduce the exact mix chosen by the user on the client node. The MixedMessage may then be stored as a .wav file in the database on the host. It is contemplated that the actual mix process, the storage of the file, additional audio processing during the mix, output file conversion to different formats may take place on the host node.

of the present invention, the MixedMessage can be uploaded to a voice mail system

chosen by the user for use as a voice mail greeting. The MixedMessage may also be used to create an "intentional message" to be sent to a recipient's voice mailbox. In yet another aspect of the present invention, the MixedMessage may be utilized to create an email message which includes the MixedMessage along with associated text and the

5 chosen image.

It further contemplated that during the recording and creation of the MixedMessage, additional processes may take place, also. For example, it is contemplated that the user may manually modify the contours of the interleaving process. This modification may be accomplished by presenting the user with controls presented during the mixing phase which may be used to adjust the level and time contours, for example.

Furthermore, it is contemplated that the user may also modify the recorded message with processing tools to enhance or modify the voice information. For example, the user may be able to make their voice sound similar to that of a popular character or

celebrity. It is contemplated that this procedure may be accomplished by presenting the user with audio processing tools during the recording or mixing/audition phases.

While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications

5 than mentioned above are possible without departing from the inventive concepts herein.

The invention, therefore, is not to be restricted except in the spirit of the appended claims.

09630920-100600